IN-82
24113
40P

NASA
Technical
Memorandum

NASA TM-108438

A SURVEY OF GEOGRAPHICAL INFORMATION SYSTEMS APPLICATIONS FOR THE EARTH SCIENCE AND APPLICATIONS DIVISION, SPACE SCIENCES LABORATORY, MARSHALL SPACE FLIGHT CENTER

By D. Rickman, K. A. Butler, and C. A. Laymon

Space Sciences Laboratory
Science and Engineering Directorate

February 1994

(NASA-TM-108438) A SURVEY OF GEOGRAPHICAL INFORMATION SYSTEMS APPLICATIONS FOR THE EARTH SCIENCE AND APPLICATIONS DIVISION, SPACE SCIENCES LABORATORY, MARSHALL SPACE FLIGHT CENTER (NASA. Marshall Space Flight Center) 40 p

N94-28820

Unclas

G3/82 0002413

National Aeronautics and
Space Administration

George C. Marshall Space Flight Center

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1994	Technical Memorandum

4. TITLE AND SUBTITLE

A Survey of Geographical Information Systems Applications for the Earth Science and Applications Division, Space Sciences Laboratory, Marshall Space Flight Center

5. FUNDING NUMBERS

6. AUTHOR(S)

D. Rickman, K. A. Butler,* and C. A. Laymon**

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

George C. Marshall Space Flight Center
Marshall Space Flight Center, AL 35812

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, D.C. 20546

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NASA TM- 108438

11. SUPPLEMENTARY NOTES

* New Technology, Inc., Huntsville, AL
**Institute for Global Change Research and Education, Huntsville, AL

Prepared by Space Sciences Laboratory, Science & Engineering Directorate.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified--Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The purpose of this document is to introduce Geographical Information System (GIS) terminology and summarize interviews conducted with scientists in the Earth Science and Applications Division (ESAD). There is a growing need in ESAD for GIS technology. With many different data sources available to the scientists comes the need to be able to process and view these data in an efficient manner. Since most of these data are stored in vastly different formats, specialized software and hardware are needed. Several ESAD scientists have been using a GIS, specifically the Man-computer Interactive Data Access System (McIDAS). McIDAS can solve many of the research problems that arise, but there are areas of research that need more powerful tools; one such example is the multispectral image analysis which is described in this document. Given the strong need for GIS in ESAD, we recommend that a requirements analysis and implementation plan be developed using this document as a basis for further investigation.

14. SUBJECT TERMS	15. NUMBER OF PAGES	16. PRICE CODE
Geographical Information System (GIS), Requirements, Survey, Introduction, MSFC	43	NTIS

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

ACKNOWLEDGMENTS

This report originated in the needs of the MSFC/Earth Science and Applications Division. It has benefitted from the input and review of many people. Members of the Division from whom reviews were received include Briscoe Stephens, Jeff Luvall, Paul Meyer, Gary Tom, Karen Parker, Gary Jedlovec, and Dale Quattrochi. Additional reviews were provided by David Walter of NASA/Stennis Space Center, Mary Viv Lawson of PCI, Tom Smith of Intergraph, Ferron Risinger of Delta Data Systems, and Bernard Kuennecke, Chairman, Department of Geography, Radford University.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

TECHNICAL MEMORANDUM

Geographical Information Systems Applications for the Earth Science and Applications Division Space Sciences Laboratory Marshall Space Flight Center

1.0 INTRODUCTION

Much of the research within the NASA Marshall Space Flight Center's (MSFC) Earth Science and Applications Division (ESAD) is interdisciplinary in nature. Scientists with backgrounds in meteorology, hydrology, geology, and numerical modeling all work together with the common goal of understanding the Earth's atmosphere. The geographic data sets these scientists handle are best managed and analyzed within the domain of computer software known as Geographical Information Systems.

A Geographical Information System (GIS) contains software that operates on data which is linked to the globe. It is used to obtain and analyze geographically-related information. Uses of GIS technology are extremely varied. For example, one could be interested in the effect of the spatial variability of vegetation on mesoscale hydrologic budgets; the relationship between microwave signature, emissivity, and soil moisture; the effect of winter storm fronts on the transport of sediment along the Gulf Coast; or simply require a tool to geo-register aircraft remote sensing data to the ground. All of these activities involve processing data inherently linked to the globe; and GIS software is the class of tool used to perform them.

From the examples of GIS applications presented above, clearly most of the research conducted within ESAD is within the domain of GIS. In fact, several scientists have independently sought hardware and software solutions to meet their GIS needs. Others are currently seeking to fill additional needs and are evaluating several different GIS systems. Consequently, there has been little coordination of effort or solutions.

A survey of current Division-wide research and GIS applications was conducted to identify the status of GIS use. The findings of this survey are contained herein. Every scientist surveyed uses large quantities of raster-based data which are tied to the globe in some manner. They analyze these data in a wide variety of ways, but the methods, for the most part, are well within the realm of existing GIS systems. Division researchers are already using several GIS systems. In fact the primary data analysis tool used by the Division, the Man-computer Interactive Data Access System (McIDAS), is a GIS. It is, however, a tool that is nearly a decade out of date in many critical aspects.

On the basis of this survey, it is certain that much of the current and future research would benefit from more effective use of current GIS technology. A broader evaluation of GIS requirements and development of an implementation plan for the Division as a whole would streamline support services and ensure that resources are allocated efficiently to meet current and future research requirements.

This document begins with an overview of GIS terminology which includes two examples of Division GIS use. This is followed by a discussion of the GIS related needs of the Division based on our discussions with the scientists. Sections on limitations and requirements are included and should be used as the foundation for a further requirements analysis and implementation plan. Finally, to help anyone not familiar with GIS technology we have included several appendices. These are: a brief glossary, a reference list, and list of a few vendors with some price data.

2.0 GIS OVERVIEW

2.1. GIS Definition

A geographical information system is an integrated set of utilities for the collection, storage, and analysis of geographically referenced data (figure 1). "The key features which differentiate GIS from other information systems are the general focus on spatial entities and relationships, together with specific attention to spatial analytical and modelling operations. In a technical sense it is the ability to organize and integrate apparently disparate data sets together by geography which makes GIS so powerful." (Maguire et al. 1991). Thus, GIS incorporates features of other information systems, such as remote sensing, data base management, and computer cartography with the addition of a geographic component. Every GIS contains the following elements (derived from Star and Estes,1990):

A. Data Acquisition - gathering and storing data derived from sources such as maps,remote sensors, aerial photography, Global Positioning Systems (GPS), data loggers, etc. Data of many disparate types from different sources are commonly needed by users of GIS systems.

B. Preprocessing - reformats data for use with the GIS, involves data structure (raster or vector) and data media (tape, digitized data, diskette, logger memory) conversions.

C. Data management - creation and query of the database itself, i.e., data entry, update, delete, and retrieve.

D. Manipulation and analysis - analytic operations that manipulate image and database contents to derive new information.

E. Product generation - output statistical reports, maps, graphics and animations.

2.2. Data Types

There are three dominant types of data used within GIS packages: raster, vector, and tabular. Nearly all packages are designed to primarily handle only one of the three data types while the others receive subordinate attention.

2.2.1. Raster Data

Raster data are arrays of x,y locations, which are referred to as pixels (picture element), where every point in the array has a value. This data type is well suited to the storage of imagery and continuous surfaces, such as topography or precipitation fields. Most raster-based GISs are oriented toward processing remotely-sensed data commonly from satellites. Therefore, their data structures and data handling tend to be efficient at handling arrays large in two dimensions, x and y. A third dimension, used for the multiple spectral bands, is usually limited in range. There is rarely any explicit provision for more than three array dimensions. Because of the nature of remotely sensed data, the data are usually limited to 8-bits per pixel. Only the more advanced packages allow data storage and manipulation of more than 8-bit integer data.

2.2.2. Vector Data

Vector data are stored as strings of x,y pairs representing the geographical position of points and lines. Almost always these strings have associated "attributes" to describe what the point, line, or shape represents. A collection of vectors making a contiguous, closed feature is frequently referred to as a polygon. The vector is an efficient data storage format when the data are either linear features (streams), boundaries (contact between air masses), or regions (soil types or vegetation classes). There are a wide range of vector storage formats (see Appendix B), some of which are extremely complex.

2.2.3. Tabular Data

Tabular data are attribute information that are stored in a table format of a database. These are intrinsically linked to geographic elements. Examples include radiosonde telemetry, population statistics, economic information, rainfall records. GIS software oriented toward such data is used by libraries, businesses, and national and international agencies.

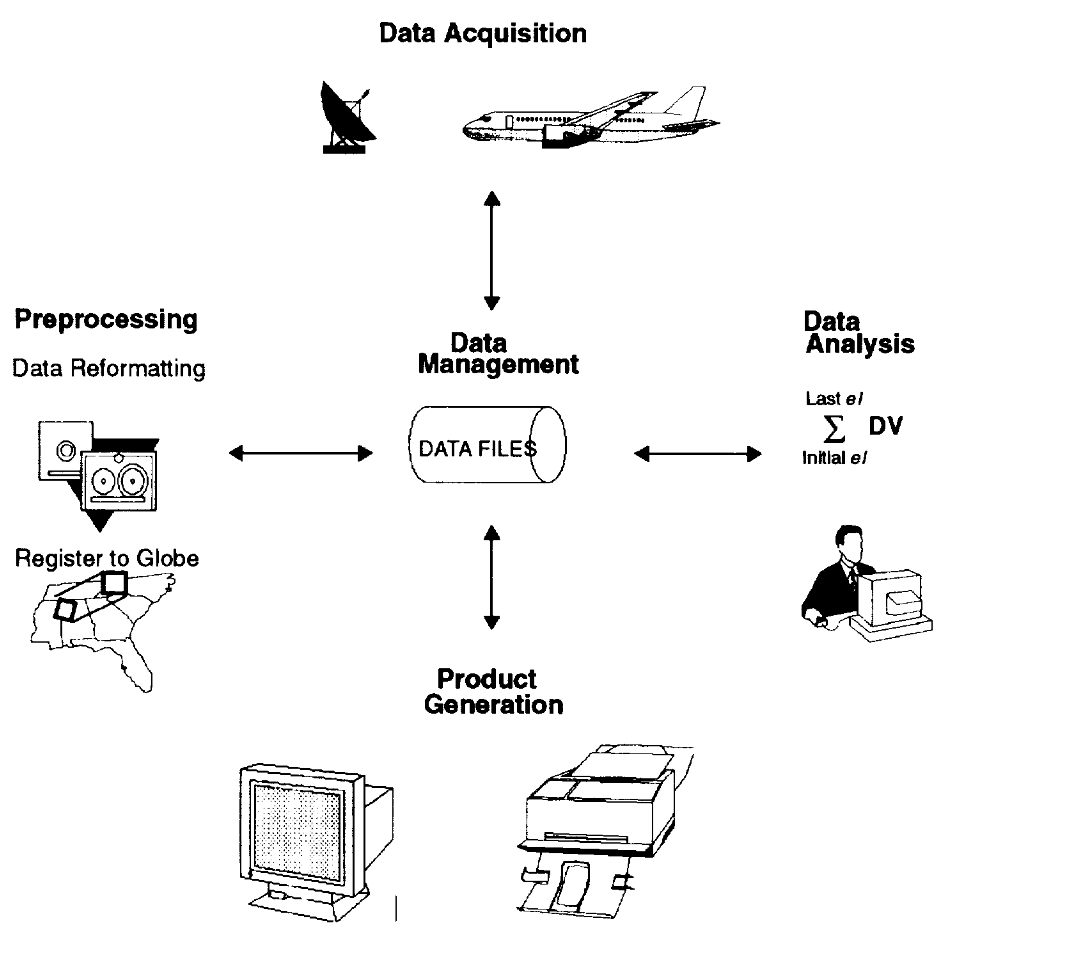

figure 1. All operations use a common format and communicate to each other through the common format.

2.3. Example Division GIS Cases

2.3.1. The Convection and Precipitation/Electrification Hydrometeorology Project

The Convection and Precipitation/Electrification (CaPE) Hydrometeorology Project (CHymP) is one example of the Division's need for GIS technology. It illustrates the way GIS technology can be used to process data from numerous disparate sources in preparation for modeling. The Intergraph Corporation's InterPro 6487 Image Station and its MGE and ISI families of software are being evaluated through this project.

The objective of this project is to model in three dimensions the land and atmosphere water and energy budgets on a daily time scale. In order to accomplish this objective, estimates of the various components of the land and atmospheric water and energy budgets are required for model input and validation. These data were obtained using in situ surface and atmospheric measurements, satellite and aircraft remote sensing imagery, and geographic data. In so doing, the project brings together many data sets of various types, formats, sources, and spatial and temporal resolutions (see Table 1), all of which must be synthesized using a GIS with a wide range of conversion utilities and processing functions. Current GIS technology requires modeling to be performed offline. Much of the requisite model input data are preprocessed with the GIS and output in gridded format thereby making the data compatible for modeling. Model output will also be in gridded format and re-ingested into the GIS for display and analysis.

Requirements:

The CHymP requirements call for an information system with a broad range of functionality. The ability to import data from a wide variety of disparate sources and from a wide variety of formats is perhaps the most overlooked aspect of information systems, and yet is one of the most important to interdisciplinary science today. It is necessary to be able to import data from a variety of software packages such as McIDAS, ARC/INFO, ELAS, etc. Likewise, it is necessary to be able to ingest data from a variety of "standardized" formats, such as ASCII and DLG/3-Optional. These import utilities must be flexible enough to accept variations in the standard formats without impeding progress. The system must be able to easily import data from the SPOT, Landsat TM, and AVHRR satellites, as well as non-georegistered data from aircraft-based video and multispectral instruments.

Once data are ingested by the system, it must be possible to modify map projections and coordinate systems. Simple display functions are required, such as displaying both raster and vector data together and placing text on imagery to label features or sites. It should be easy to compute information about vector elements such as stream lengths and polygon areas. Vector elements should be linked to attribute tables in a common database format or software product, such as Informix or Oracle.

Once image data are imported into the system, it must be easy to co-register one image to another or rectify the image to a map. Investigators desire a utility to navigate less common satellite data for which a platform-specific import utility doesn't exist. Investigators must be able to easily perform image classifications and query the contents of individual pixels. They must also be able to co-process images with different resolutions without resampling one of the images. They must be able to collect statistics about the image contents based on a rectangular subarea or irregular polygon. These statistical results must be able to be output in ASCII format.

Additionally, it would be highly beneficial to be able to collect histograms in a batch mode on multiple images. Investigators would like to be able to output raster image data selected by a fence or polygon in an ASCII array format to study the pixel contents of the specified region.

Integrating vector and raster data must be performed without converting the vector data to raster. For a typical example, see figure 2. It must be possible to extract information from an image or raster dataset based on vector elements. An example of this is to determine the rainfall amount in a specified watershed where the rainfall is a gridded dataset and the watershed is identified by an irregular vector polygon. Scientists must also be able to view the raster data as a 3-D surface and project vector elements on the surface.

Thus far, GIS has served as a preprocessor for data prior to environmental modeling. Ultimately, investigators need to be able to link environmental models directly with GIS software so that the GIS can facilitate visualization of the models variables in real time. Similarly, the package must be linked to graphing capabilities so that observed versus modeled results can be visualized in real time as well.

Outputting and exporting data is as critical to a workflow as any other process. A GIS is useless without the ability to output results for presentation. It must be able to produce finished maps including legends, coordinate system grids, north arrows, and scale bars. Investigators require the ability to dump the screen to a printer to document and annotate work in progress. Similarly, it should be easy to capture work displayed on the CRT and output the whole screen or a portion of it in any of a number of formats, such as RGB, Tiff, and PCX.

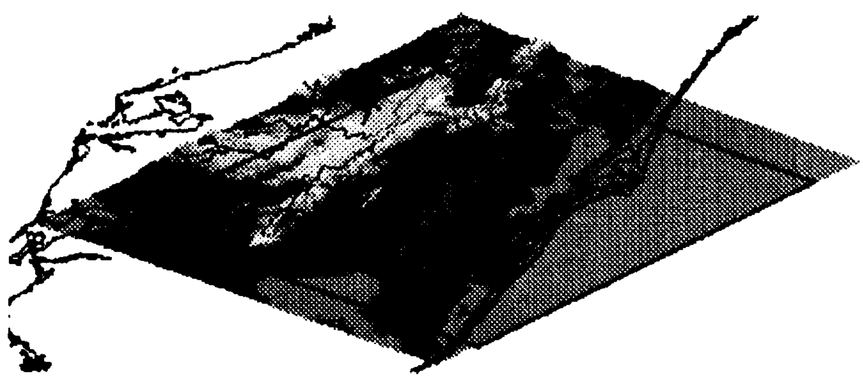

figure 2. An isometric view of the Cape Hydrometeorology Projects research domain in central Florida showing a raster dataset of color-coded topography with vector data projected onto the surface delineating the coastline and major watersheds.

TABLE 1. CaPE Hydrometeorology Project (CHymP) Data Sets*

Variable	Source	Data Type	Medium
Precipitation			
raingages (212)	various	Point	various
WSI composite (5 WSR-57 radars)	WSI Inc.	Raster	MO
CP-2, CP-4 radars	NCAR	Raster	8 mm tape
Other Meteorological			
surface flux sites (7)	various	Point	
PAM II (47)	NCAR	Point	8 mm tape
Florida Dept. of Ag. stations (5)	Florida DOA	Point	ftp
KSC wind towers (51)	KSC	Point	8 mm tape
NWS (11)	NCDC	Point	Diskettes
radiosonde sites (11)	NCAR	Point	8 mm tape
Streamflow	USGS	Point	Diskette
Satellite			
SPOT (3 images)	SPOT Image Corp.	Raster	9-track tape
GOES VIS, IR	NCAR	Raster	MO
MAMS	MSFC	Raster	
Geographic Information			
digital elevation	NGDC	Raster	9-track tape
soils	SCS	Vector	9-track tape
basin delineations	USGS	Vector	9-track tape
hydrography	USGS	Vector	9-track tape
land cover/vegetation type	FL DOT,NGDC	Raster	ftp

*Provided by Bill Crosson

2.3.2. Multispectral Image Analysis and Product Interpretation for FIFE

A second example of GIS use in the Division is the First International Satellite Land Surface Climatology Project (ISLSCP) Field Experiment (FIFE). Multispectral image data from the airborne Multispectral Atmospheric Mapping Sensor (MAMS) was collected at multiple times and on multiple days conducted during FIFE 1987 (see figure 2). Image data was collected in 7 visible and 3 infrared channels during a total of eight overpasses of the FIFE study site (a 15 x 15km region south of Manhattan, Kansas). The data were navigated (earth located), calibrated (converted to radiometric quantities), and registered (remapped to a defined earth projection covering a specific region) on the MSFC mainframe McIDAS system using standard and user specific routines. Geophysical parameters (e.g., land surface temperature (LST), and Normalized Difference Vegetation Index (NDVI)) were derived from the calibrated image data to study the ability to monitor spatial and temporal variability of these parameters from future geostationary satellites.

Scientific Objectives:

A. Demonstrate the spatial and temporal variability of the derived parameters through visualization of the derived products and their input data.

B. Study the variability and relate it (in a statistical way) to:

(1) observational constraints such as view geometry, solar zenith angle, time difference of data, day/night view, etc. Note: These data are in vector form.

(2) underlying geophysical characteristics (makeup) of the surface and related surface parameters (land use, elevation, terrain, soil moisture, etc.) which come from either in situ or remote measurements. Note: this data is in both vector and raster form.

GIS equirements:

A. The ability to import 16-bit McIDAS data maintaining full data integrity is required.

B. The GIS must be able to display multiple channel/parameter/time information without loss of resolution.

C. The ability to statistically interact with the data to develop the inter-relationships among variables is needed.

D. The GIS must have a data base for both vector and raster data or the ability to import one (land use, DEM data, and a number of other surface geophysical parameters) which describe the FIFE study site. This data may include remote sensing measurements from other observation platforms (SPOT, Landsat, other aircraft scanners, etc).

E. The ability to produce color hardcopy of the input and derived data (images) and statistics (tabular data) is necessary.

The Problem:

McIDAS is well suited for the processing and calculation of derived parameters from image data but is limited in its display (visualization) of multispectral information and in its ability (actually inability) to analyze relationships between variables.

Several other GIS packages were proposed (ELAS and Linkwinds) as a solution to the limitations of the visualization and analysis capabilities of McIDAS. The two software packages proposed to address the above requirement do not fully satisfy them. Functionally they probably have the ability to address items 2,3, and 5. They fall short of requirements 1 and 4 for three reasons. Firstly, they require the development of software to transfer data from McIDAS to some other data format (which is different for each recommended package). Secondly, Linkwinds does not handle 16-bit data. And finally, the required GIS data bases in requirement 4 are non-existent for the region of interest and would need to be imported into the individual software packages. This would require specific programming and conversion of each new data base (DEM data, land use, etc) which has no common input format.

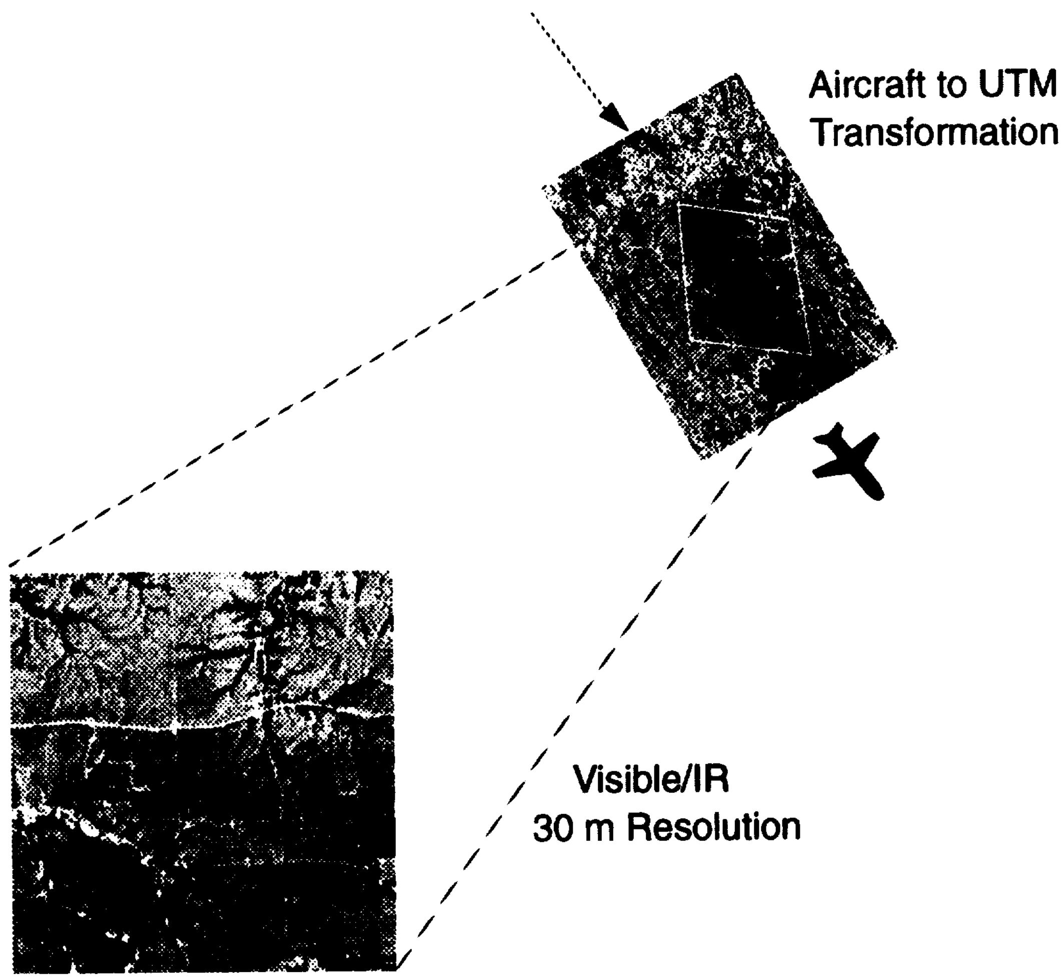

figure 3. MAMS Data for FIFE.

3.0 SUMMARY OF INTERVIEWS

There are several themes that ran through many of the discussions we had with the Division's researchers. We have summarized them below:

A. Virtually all of the work done in the Division is explicitly tied to the global coordinate system.

B. Many of the researchers are unaware that the work they are doing is well within the scope of GIS; McIDAS users are unaware that they are already using GIS software.

C. Many scientists already use or are evaluating one of several GIS packages for some part of their research - McIDAS, Intergraph, ELAS, AGIS.

D. Most of the raw data is raster. There are some data in vector and tabular form, but most of it appears to be converted into raster formats for processing.

E. Most scientists are using multiple data sets from disparate sources. With this are inherent problems in:

1. media compatibility
2. data formats - practicality of conversion (no one format is suitable for all types of data!)
3. non-congruent coverage between data set
4. different resolutions
5. different projections

F. Sometimes large quantities of data are involved; therefore, data management (storage and record keeping) is burdensome.

G. The dimensionality of the data varies considerably. Some are relatively small in x,y with very large time components. Others have x,y,z with multiple variables at several times. The highest dimensionally found in our discussions was four independent variables with one dependent variable.

H. Most of the raw data can be reasonably stored in byte format. However, there are several places where the ability to store and manipulate other formats (i.e., floating point or 32-bit integers) is critical.

I. Often repetitive tasks are needed; therefore, automation of some routines is needed.

J. Virtually all wish to do models based on the geographically related data.

K. The researchers have a clear desire and need, supported by experience, to modify existing algorithms as well as create new ones.

L. Several researchers, who have spent time evaluating one or more packages, are already aware of how much time it takes to learn a GIS.

3.1. Scientist's Concerns

Many of the researchers' needs and concerns are easily addressed. However there are several concerns that need special attention as they present particular problems.

3.1.1. Data Format Conversion

Any time files are transferred between machines or between software packages, there is usually a problem with data format conversion. It is typical to spend 1-3 days simply getting a new type of file or source of data into a user's GIS. This is an obvious waste of time and talent.

3.1.2. Training

Training and documentation is a major concern for the users of GIS technology. The amount of effort needed to become proficient with the basic concepts and operations in any package is measured in weeks and months, even in the best of cases. Additional time is required if one wishes to modify or add programs within a package. The amount of time needed to learn a GIS is sufficient to affect the planning of the researchers.

3.1.3. Dimensionality of the Data

The dimensionality of the commercial raster-based GIS software is restricted to three dimensions. Two of the dimensions are used for x,y (longitude, latitude). Mathematically these are the independent variables. The third dimension is used for the dependent variables. These may be bands from the electromagnetic spectrum, topography, or any other single variable which form complete arrays. For much of the research conducted at the Division, additional independent dimensions are necessary. These would be used to hold altitude, elevation and time.

3.1.4. Animation

Generation of an animation is a major desire for several of the scientists. This has generally been at the fringe of GIS technology. It is also highly dependent on hardware. Therefore, a large number of GIS packages have little or no animation capability. Those that do, offer the user relatively little control or flexibility.

3.1.5. Data Integration

Many of the researchers need to integrate image data from multiple sources. Almost invariably this requires the matching of different pixel sizes as mapped on the ground. The current technology requires the user to resample all data sets to a single resolution, forcing either an expansion of storage requirements or loss of information.

3.1.6. Modeling

Modeling is the ultimate goal of most of the researchers in the Division. There is no generic ability to create models in the sense the scientists use. Many of the vendors can correctly claim modeling ability, but their concept of a model is different. In fact, the vendor community uses several concepts when they refer to "modeling." It is very important that we carefully define our terms when we speak to the vendors on this subject.

3.1.7. Data Storage and Network

The choice of data structure is usually controlled by the source of data. Those working with remotely-sensed imagery use raster-based systems. Those who must digitize maps or networks (roads/streets, sewer systems, stream drainages) will use vector systems. It is possible for each of these data structures, raster, vector, and tabular to be used to express all types of data. However, to do so is usually inefficient in terms of storage volume, access time and rate of manipulation. In practice, the type of data storage is critical, with effects at many levels. The exact details of storage, all the way to the grouping of data onto a hard disk's surface, ultimately control what types of operations are practical within a given time. This is because the quantity of data involved is so large. For example a single satellite image from a common source, the Landsat Thematic Mapper (TM), contains almost 300 MBytes of data. It is stored as a three-dimensional array several thousand units long on two axes and seven units long on the third axis. Even with a good disk (600 kiloBytes per second average transfer rate) it takes approximately 500 seconds to read every value in such a file. Any inefficiency in data access is replicated by the number of accesses. Raster-based systems are particularly sensitive to this problem. Typical projects in such systems can often have a GByte or more of basic data files. Vector-based systems usually have 1%-10% of the data of a typical raster system. This is still a large quantity when compared to what most software handles on a routine basis.[1] The large quantities of data also limit the utility of file server configurations. Typically ethernet-based systems run at 25-500 kiloBytes per second, average file transfer rate. The high number is unusual; the lower value is common. At the lower rate it would take a user over 3 hours to read the 300 MByte TM file mentioned above just once. Unfortunately with current technology, networks are simply too slow to permit much use of a server for raster-based operations. Networks are much more practical for vector-based systems, but even there the GIS work should be isolated from other users by a bridge.

[1] Large word processing files, such as this document, are measured in tens of kiloBytes.

4.0 LIMITATIONS

4.1. Hardware Limitations

4.1.1. On-Line Disk Storage

Geographical information systems are very complex and require a large amount of on-line disk storage, typically greater than 100 MBytes. Although, it is possible to store the software on a server, there is likely to be a significant performance degradation for the users. This is due to the size of the individual executable components of the GIS package and need of users to swap in and out of modules frequently. The normal solution is to provide at least a 1-GByte disk locally, for software and temporary data storage. Additional on-line storage is common.

4.1.2. Network Speed

Given the nature of the work being done here, efficient utilization of the technology requires ready access to very large data storage capacity at each "seat" (a seat being where a user works with the GIS software, and by implication the associated hardware and software). This must be dismountable, random access, read/write media. This is needed because the quantity of data used and produced by a modern GIS package is measured in the hundreds of MBytes. It is not practical or cost effective to maintain all of this information either locally or on a server. The second deficiency is the bandwidth of the existing network. If multiple people wish to work over the network, using a 10 Mbit/second bandwidth is too small.

4.1.3. I/O Bottleneck

It has often been suggested that one could move a raster-based GIS package onto a super computer, and it has been done. It gains the user little advantage because the CPU is not the limiting factor in most cases. Due to file size most raster-based GIS tasks are I/O bound and disk transfer rate is the limiting factor. Therefore, hardware configurations optimized for GIS applications have large capacity (several GBytes), high-speed disks with good continuous access rates and one or more read/write dismountable, random access media. A 1-GByte drive with good random access is also desired to hold the GIS software itself, as well as the operating system and other applications. The former are used for data being used immediately; the dismountable media are for most of a project's active data. Tape is used only for archives.

4.1.4. Display Technology

Display technology is also critical. In a typical color computer display the user is limited to 256 possible simultaneous colors. These are termed 8-bit displays. Most GIS packages will use 24-bit displays to very good advantage. Imagery with more than one dependent variable, often termed either bands or channels, such as the AVHRR or MAMS, is mucheasier to interpret when displayed with the 24-bit display.

4.2. Software Limitations

4.2.1. Handling of Data Types

Most software packages are efficient at handling only one of the three types of data: raster, vector, or tabular. At best, functionality will be poor or insignificant with the other data types. This is almost a universal situation. It is widely recognized that most users experience the problem, and vendors are trying to solve it. Because of the difficulty in doing so and the wide-spread need, many different approaches have been used. Most are cumbersome and inefficient. Anyone reviewing a software's suitability for a task must be extremely specific when questioning how it works when handling "other data types."

4.2.2. Resolution Problem

In general, to integrate two different data sets requires that both sets be in the same map projection and at the same resolution. For example, if one has AVHRR data at 1 km and aircraft imagery at 5 m, one must either resample the AVHRR to 5 m or the aircraft to 1 km. Therefore, either the data storage requirement increases by a factor of 40,000 or most of the information inherent in the aircraft source is eliminated. A solution to this problem is being investigated by Delta Data Systems, under a Phase 2 SBIR with NASA.

4.2.3. Functionality vs. Complexity

There tends to be an inverse relationship between the software's power/functionality and the complexity for the user. This relates to the time it takes to create a large package and the evolution of the computer industry. Major GIS systems, commercial and public domain, are almost all second-generation software.[2] Some, especially the commercial packages, have a veneer of a graphical user interface (GUI) on top of a command line-driven package. But, due to limitations of converting a command line structured architecture into something else, the primary power is usually obtainable only through the command line. Few GIS packages are really third generation, and we are aware of none that are fourth generation.

Newer systems, incorporating either native graphical user interfaces or object-oriented and multitasking paradigms, are too new to have the full panoply of tools that the older packages have. The major packages represent many hundreds of **manyears** of programming effort. Their strength comes from the number of explicit functions or operations programmed into them, or the ability to generate extremely complex processing flows.

---

[2] generations of software: 1st - Minimal memory, support utilities, extremely expensive, CPU vs. terminal. Resulted in absolute minimum support for the user. Thus, a high degree of skill or knowledge required by the user. 2nd - Reduction in memory cost allowing larger programs. Part of this increased memory was used to upgrade the user interface, becoming more English oriented, with on-line help and prompts. Also, the concepts of modularity and interprocess communication. 3rd - User interfaces are graphically oriented, processing is desktop. 4th - Application software is object-oriented and hardware/OS are multi-tasking.

Also note that the adoption of a GUI is not the total solution. There are two serious problems when adopting a GUI interface versus the command line approach. The GUI is easier to learn, but it is more difficult to use for repetitive tasks. Repetitive tasks arise frequently in research. For example, take data from more than one timestep, perform the same operations on the imagery, and compare the differences. The other problem with a GUI is, for the experienced user, that it is slower. It is much quicker to type a single command line than it is to descend through a string of icons, windows, or menus. Because of these problems, the best new software systems will have both a GUI and a command line interface designed into them from the start.

4.2.4. Learning Curve

Corollary to the power/complexity relationship is the time that must be invested to learn a major package. Two of the Division's scientists have spent approximately 6 months each learning only a small portion of one major package. This is fairly typical. If one is to profit from a GIS, a significant effort must be given to learning how to use the software. The amount of time and effort varies enormously from package to package and is dependent on the level of expertise needed. It is also strongly affected by the availability or absence of some kind of consultative expert. If an expert is available, it would greatly reduce the level of proficiency needed by the individual user.

Having learned a particular package, for most users it is extremely difficult to switch to another package. For comparison, think of the difficulty of learning a new word processor. All word processors do essentially the same things, yet none are operationally interchangeable. The more powerful GIS packages are many times more complex than any word processor, and switching between GIS packages can be correspondingly difficult. There are two reasons for this. First, each developer has their own interface design. Second, the fundamental concepts behind each package are often quite different.

4.2.5. Flexibility vs. Ease of Use

Another problem with existing GIS technology is specifically related to its use in a research environment. For most software, the user is unable to do anything other than what is expressly intended and permitted by the developers. There is a set processing flow and all data will go through the predetermined paths. In commercial systems, the user is also carefully shielded from intricacies and algorithm details. While this is reasonable in a word processor or spread sheet, it becomes a problem when a researcher uses a powerful GIS.

Large GIS packages are incredibly powerful. This power comes from two sources: 1.the number of functions directly and explicitly available, and 2. the freedom to use the functions in virtually any order. Indeed, a large GIS can be considered a programming "language" consisting of high level operators which the user can string together in any desired order. As with any computer language, the number of ways to accomplish a task is essentially limitless. It is this flexibility that makes a GIS particularly useful in a research environment.

However, when writing any software it is easy and normal to check and prevent certain situations, either by blocking a processing path or by preventing access to process-related variables. This is often done to protect the user from accidental errors. The problem comes from deciding what is an appropriate situation. Compared to most GIS users, researchers frequently have a much better understanding of the concepts involved and can, therefore, push the envelope safely. Second, researchers must always push to and beyond the limits of standard processes. They will, therefore, frequently need to exceed built-in limitations to protect ordinary users. The researcher's needs and abilities will clash with the programmer's need to protect the user from mistakes. There is no clear way to resolve this conflict.

Tied up with this problem of flexibility is the topic of "ease of use." To simplify what the user must know to function, software developers frequently use a large number of processing checks to catch accidental errors. These checks have the side effect of decreasing flexibility. We observe that to some extent the easier the GIS software is to use, the more limited is its flexibility.

Therefore, when selecting a GIS the related problems of flexibility and ease of use must be considered. The user must anticipate in advance how often the standard processing flows in a package will or will not be acceptable. As this is not practical in most cases, one of two approaches can be used. First, learn the type of work for which a package is being used and assume that the software can accommodate that type of work. Second, find someone whose work is similar to your own and ask how his software succeeds and fails to satisfy him. Either approach will help; neither is completely successful.

The needs of the Division's researchers for flexible control of processing will task any package. Although intangible and difficult to quantify, we should carefully consider flexibility vs. ease of use when looking at GIS software.

4.2.6. Data Format

There are several efforts now underway to address the data format problem. The USGS has helped create a data transfer standard, the Spatial Data Transfer Standard (SDTS). This is now a Federal Information Processing Standard (FIPS), and is contained in FIPS 173. It covers both raster and vector data types. Ultimately, this will help standardize the translation of data between disparate software packages. We need to insist that any GIS software we purchase support and use this standard.

NASA's EOS program has adopted the Hierarchical Data Format (HDF). This format is supposed to produce a file structure that is identical at the binary level on all machines. This means that a user on any machine can access the data without concern for machine-specific details regarding bit significance, or byte and nibble ordering and floating point storage specifications. Unfortunately, the format as it stands does not have specific structures with definitions for information essential to GIS work. For example, the size of a pixel and the projection of the data are not defined entities.

There are several efforts to create generic reformatting routines. These are being produced by hardware vendors, software vendors, and the U.S. Government. The generic tools require one to know the **exact** details of both the incoming format and the outgoing format. An understanding of data format, blocking, sectors, headers, floating point structures, and other details affecting data storage is usually required. These generic routines are not flexible enough when translating the more complex formats.

Another generic approach is also being used. Here the program contains the details for a large number of formats. The user can then come from any of the supported formats and go out to another supported format. The problem is the enormous number of formats that exist.

The course of last resort is to write a reformatter specific to the individual's needs. This usually requires someone who is at least moderately skilled in the GIS being used, as it will usually require use of the subroutines already in the package.

4.2.7. Data Management

It is common for a researcher using a GIS to create dozens of files in a single session. A project that lasts any significant length of time will quickly be burdened with tracking all of the files. There is no system of which the authors are aware which does this tracking automatically.

4.3. Facilities Limitations

The Division's GIS will require expansion and/or additional facilities. Shared work stations with a common input/output area to support the digitizers, scanners and plotters are desirable to optimize utilization of the total GIS system.

Scientists will require the assistance of technicians to administer the GIS hardware, software, and data to ensure timely scientific processing.

Physical layout and equipment placement must incorporate GIS ergonomics and requirements during the planning stage.

5.0 REQUIREMENTS

Given the research being performed at the Division, having discussed GIS issues with the scientists, and having reviewed recommendations for GIS functions and utilities needed for global applications suggested by Clark et al. (1991), we have derived general requirements, summarized as follows:

5.1. Data Acquisition and Preprocessing

A. The GIS needs to be raster-based with the ability to integrate vector data. Vector-to-raster and raster-to-vector conversions are also required.

B. Conversions are needed for all major data formats, both input and output, such as McIDAS format, Digital Line Graph Structure (DLG), Hierarchical Data Format (HDF), Spatial Data Transfer Standard (SDTS), and Landsat. And if a new format is required, software and/or vendor support must be available to assist in developing conversion software.

C. The GIS must be able to handle 8-bit (Byte), 10-bit, and 24-bit integers and real data values in ASCII or binary forms with conversion and editing capabilities.

D. Multi-dimensional data (x, y, z, time, parameters) need to be handled in an efficient manner.

5.2. Data Management

A. The GIS must be integrated with a state-of-the-art Data Base Management System (DBMS).

B. The ability to edit vector and raster data using DBMS/spreadsheet algorithm is required.

C. Conversions to/from other DBMS, i.e., ORACLE, Informix, Dbase IV, etc., are needed.

5.3. Manipulation and Analysis

A. The ability to perform analyses, contouring, and visualizations in different coordinate projections, such as Mercator and Azimuthal, is necessary.

B. The GIS must be able to convert random or gridded data to different grids and process different scales.

C. Integration with mathematical and statistical systems is a must.

D. The user needs to be able to interpolate or re-sample gridded data.

E. GIS standard functions such as overlay, merge, and statistical reports will be required.

F. There must be an ability to create and modify existing algorithms and perform modeling.

5.4. Product Generation

A. The GIS must be integrated with a fully functional image processing system.

B. The user needs to have access to pattern shading and fill, attribute placements, contour labeling, legends, annotation and scaling.

C. The user must be able to link images related to a time sequence and step through the loop manually and automatically.

D. The user will need to generate quality hard copies for publication and videos of animations.

5.5. Training

A. On-line help should be available along with tutorials which introduce basic functions.

B. User and reference documentation will be needed.

C. Courses in programming and/or usage of the systems should be available.

D. Documentation describing algorithms within the GIS are required so that the scientists know the details of how their data are manipulated.

6.0 SUMMARY

Underlaying any productive discussion is a set of common understandings. Terms, scope, and significance must be recognized by all parties before useful discourse. This document provides such a basis for GIS-related discussion. It introduces Geographical Information System terminology and explains the use and limitations of the technology. We extend this with a survey of the GIS-related work done within the organization, including summaries of interviews conducted with Division scientists.

We found that most of the research within the organization uses or even depends on computer-based geographic information systems. Indeed, the most common research-oriented software tool, McIDAS, is a GIS. The difficulty is that the technology behind this principal tool is 10 to 20 years old, and its basic utility cannot be extended in a cost-effective manner. Newer technologies are vastly more powerful and inherently more flexible. And what is especially critical for research, they are more extensible.

Further, we found that there is a growing need within our community for GIS technology. The questions being pursued require more capability in the basic software. Examples are the analysis of multispectral imagery and the integration of the many different data sources used. These are both at or beyond the practical limits of the common tool. As a result, several of the scientists have been using other GIS packages. With this search for capability comes disparity and confusion. Recognizing this as a problem, the scientists seek some common path.

It is very clear that the scientists must have better GIS tools if they are to remain scientifically competitive! Fortunately, there are many appropriate commercial packages available. However, as with McIDAS, their use is not achieved by simply purchasing the software and throwing the users in, sink or swim. Geographical Information Systems are large and complex and, therefore, have a steep learning curve for the users. Having personnel that are knowledgable with the software is almost a fundamental requirement. The packages are expensive and can represent a significant capital outlay. Finally, there are several basic types of GIS packages, not all of which are suitable for our needs.

The fact that McIDAS is used and supported demonstrates that the resources are available to procure, use, and support a GIS. We strongly recommend that a requirements analysis and implementation plan be developed to obtain the appropriate GIS.

APPENDIX A - Meeting Notes

The following are partial transcripts of notes taken during meetings with members of the Earth Science and Applications Division (ES41). They are included for completeness and are intended for reference only. The first entry in each case is from K.B.'s notes. The entry following the underscore line is from D.R.

Atmospheric Dynamics Group

The meeting was attended by Mike Newchurch of the Atmospheric Dynamics Group team (ES42) and Doug Rickman and Karen Butler of the VAT Team (ES44).

Chemical constituents in the atmosphere are stored in three-dimensional data sets with on the order of 100 data values per point. The data are derived from models, satellite, shuttle, and ground-based systems. The data are mostly in raster format with possibly some vector data. The major requirement is displaying data in a useful form, mainly at each grid point and in a specific area on the globe. There is also a need to process the data over for a time series.

Dr. Newchurch has used ERDAS and ARCINFO and feels that these products (or products like them) would not meet the group's needs.

Another major function of this group is in the area of numerical modeling of both the atmosphere and of laboratory experiments, including the Geophysical Fluid Flow Cell Experiment of Spacelab. These modeling needs are similar to the Earth System Dynamics Group's needs. The data are raster and have a high dimensionality.

Raster data with some point data. Dimensionality of the data will be a problem. He has x,y,z plus time plus many chemical species. Variable resolution also exists.

Microwave Measurement Group

The meeting was attended by Roy Spencer and Andy Millman of the Microwave Measurement Group team (ES43) and Doug Rickman and Karen Butler of the VAT Team (ES44).

The main effort of Dr. Spencer's team lies in algorithm development based on microwave remotely sensed data from satellite and aircraft sources. The scientists work with the data in two ways: first by analyzing a (sometimes very long) series of images and second by performing a case-by-case study of various data sources merged together.

The data are in raster format, with varying resolution (15-70 km). The data can have overlapping regions and areas of missing data. There is a need to easily change the mapping projection used in viewing the data.

A main concern of the scientists is that the system be easy to use, at least for performing simple functions. The imaging capabilities of McIDAS have been suitable for their use, except for the number of images allowed in a loop. The only data processing between different data sources (other than display) has been accomplished through batch programs.

Roy Spencer: Not aware that what he does use or needs GIS technology. McIDAS is a GIS. He works with misc. scanners at various resolutions, has a global scope in many cases. Does a lot of batch processing, would like to be able to interactively modify a parameter in a complex process which builds/ affects a stream of images, i.e., 15 years of daily satellite images, and then visualize what that change does to data in video mode. He works with misc. projections. Data sets are minimal in two dimensions, very large in time. Customization of algorithms is very important!! Reasonably satisfied with McIDAS. He does work with some point data.

Integrated Process Studies Group

Attendees were: Chip Laymon, Bill Crosson, Jeff Luvall, Dale Quattrochi, Ravikumar Raghaven of the Integrated Process Studies Group (ES42) and Doug Rickman and Karen Butler VAT Team (ES44).The group has some experience with several GIS systems (AGIS, Intergraph, ARCINFO). The CaPE (Convection and Precipitation/Electrification) program has many different data sources which consist of raster, vector, and tabular data. Raster type comes from radar, aircraft and satellite, and land-cover sources. Vector type is in the form of maps and soil data. The data can be of differing resolutions and formats. There is a need by the scientist to perform operations on "merged" data which goes beyond visual overlay of the data. Hydrologic models are being developed which require the use of several different data types simultaneously. The main concern of the scientist is that the GIS systems are difficult to learn and use. The group has invested time in learning the Integraph system and has reported that the learning curve for such a system is steep. The main problem that the scientists is having is that it is difficult to input the data into a system. There are many different types of formats that are common among geographical data sets. A major concern when choosing a GIS is that necessary input and output formats are supported by the system. It was suggested that the CaPE project be a sample program for determining the requirements for a GIS, because of its complexity and use of many data sources.

Mainly discussed CAPE and their experiences with it and the Intergraph. They emphasize that they need support for data ingest/reformatting. Would like to be able to model inside of the GIS, otherwise must be able to import/export data readily. 2 Max data 100 km x 30 m x 10 ch x 1/hour. Many types of data, frequently w/o complete coverage for study area. Uses raster, vector, point data. Data management is a practical problem. Multiple tables per polygon is some cases.

Infrared Measurements and Modeling Group

The meeting was attended by Gary Jedlovec of the Infrared Measurements and Modeling Group (ES43) and Doug Rickman and Karen Butler of the VAT Team (ES44).

The data is in raster format and will have several different sources and resolutions. MAMS data will be used as part of the CaPE program.

The scientist's needs are mostly in the area of visualization. McIDAS is currently being used very heavily and meets most of these needs. The main concern is that staying with McIDAS may isolate the group and cause problems with data exchange.

Accessibility to the source code which allows tailoring of the software to the scientist's particular needs was discussed. Complexity and user-friendliness of the software is also a concern

---

Gary Jedlovec: Integration of MAMS data with topographic derived values, slope/aspect, with land use information.

Separate meeting with Doug Rickman

Jeff Rothermel: airborne LIDAR (wind measurement) vs. topography and land use. Problems: the several planes sampled by the LIDAR are shifted relative to each other due to forward motion of the aircraft. As per note of March 23, he also wishes to resample cells size as needed.

Engineering Applications Group

The meeting was attended by Dale Johnson of the Engineering Applications Group (ES44) and Doug Rickman and Karen Butler of the VAT Team (ES44).

There are several databases created and maintained by this group. Data from models and actual atmospheric measurements are used for shuttle launch support. The scientists are currently using MIDDS McIDAS software. The data are in tabular and raster format. There is a need for graphics in determining vertical wind profiles and for three-dimensional analysis of actual data compared to model data.

Dale is uncertain about future directions of work. Possible need to ingest large quantities of tabular data. Models (GRAM) to be made are logically raster.

Earth Systems Dynamics Group

The meeting was attended by Kevin Doty, Bill McCaul, and Bill Lapenta of the Earth System Dynamics Group (ES42) and Doug Rickman and Karen Butler of the VAT Team (ES44).

Mesoscale and global numerical models are run and their results compared to other models as well as satellite, radiosonde, and other actual datasets. The LAMPS (Limited Area Mesoscale Prediction System) and the RAMS (Regional Atmospheric Modeling System) are examples of the models which are run on the CRAY system. The results of the models are plotted using NCAR (National Center for Atmospheric Research) graphics.

Most of the data are in raster format with some tabular data (radiosonde data). The data are three dimensional in space, contain several variable for each point in space, and are also stored by time, giving a five-dimensional dataset.

The main need for a GIS is in the comparison of actual data to model data, so that the model predictions can be verified. Visualization of these data with several mapping projections is also a major requirement. The ability to create derived fields within a GIS is also desirable.

They have data with dimensionality 5, at least 4 orders of magnitude needed for precision, array sizes are small on any one dimension (in the hundreds). Use GIS to compare model results with observed data (satellite, rawinsonde, etc.) and for visualization.

APPENDIX B - Abbreviations and Definitions

AGIS - Automated Geographical Information System, product of Delta Data Systems. See vendor data, Appendix D.

Attribute data - data other than location information. This can be any kind of data of interest, such as temperature, chemical content, population, distance to something. Attribute data are usually used with vector, not raster or tabular data sets.

AVHRR - Advanced Very High Resolution Radiometer. A sensor onboard a series of satellites. The name is now a misnomer as the spatial and spectral resolution are rather low.

Byte - A unit of computer memory or storage with 256 possible values. In ASCII a byte is used to store each alpha-numeric character. In remote sensing a Byte is often used to store the intensity recorded for each pixel.

CaPE - Convection and Precipitation/Electrification

DBMS - Data Base Management System

DEM - Digital Elevation Model. Also a specific data format for elevation data available from the Unites States Geological Survey.

DLG - Digital Line Graph Structure (see vector for full description)

DMS - Desktop Mapping System, a product available from Roy Welch. See vendor data, Appendix D.

DOT - Department of Transportation

ELAS - A public domain image processing and raster-based GIS package developed by NASA. It has probably been the source of more commercial spin-off products than any other public domain package.

FIFE - First ISLSCP Field Experiment

FIPS - Federal Information Processing Standard

GIS - Geographical information system, an information system that is designed to work with data referenced by spatial or geographic coordinates. GIS is both a database system with specific capabilities for spatially referenced data as well as a set of operations for working with the data.

GByte - 10^9 bytes of information. Compare MByte and Byte.

GEOS - Geosynchronous Operational Environmental Satellites

GRASS - Geographical Resources Analysis Support System, a public domain, raster-based GIS.

GUI - Graphical User Interface. X Windows 11 and MicroSoft Windows are examples.

HDF - Hierarchical Data Format

Hierarchical, quadtree, or pyramidal data structure - Four raster units are averaged to make next higher level unit. Also called quadtree for groups of four. All data stored in data set allow for faster searches.

Information system - involves observation and collection of data, storage and analyses of data, and the use of derived information in some decision making process

ISLSCP - International Satellite Land Surface Climatology Project

KSC - Kennedy Space Center

LIDAR - Light Detection and Ranging. Used for remotely measuring wind direction and velocity.

LST - Land surface temperature

MAMS - Multi-spectral Atmospheric Mapping Sensor, an airborne imaging device.

MByte - 10^6 Bytes of information, compare GByte and Byte.

McIDAS - Man-computer Interactive Data Access System. A GIS used predominantly for meteorological work.

MSS - Multi-Spectral Scanner. A sensor on the Landsat satellites. It has four bands in the visible and near IR with a ground resolution of approximately 80 meters. It was the first publically available digital data source which provided sufficient detail to be useful to a broad range of users. As such these scanners have had a major impact on image processing and GIS technology.

NCAR - National Center for Atmospheric Research

NCDC - National Climatic Data Center

NDVI - Normalized Difference Vegetation Index

NGDC - National Geographic Data Center

NWS - National Weather Service

PAM - Portable Automated Mesonet

Pixel - Picture element. An single point in the array making up a raster data set. The term derives from the use of raster video display devices to make pictures of remotely sensed imagery.

Planimetric - the correct horizontal relationship between objects on the ground

Raster - cellular organization of spatial data. Conceptually equal to a mathematical array. Each parameter of interest must be explicitly stated for each cell in a (usually regular) array over space. The term derives from video display technology.

Rectification - manipulates a raw data set so that the spatial arrangements of objects in the data correspond to a specific geocoding system.

Registration - the process of merging multiple maps so that their features overlay properly

Remote sensing - the process of deriving information by means of systems that are not in direct contact with the objects or phenomenon of interest

RFP - Request for Proposal

SBIR - Small Business Innovative Research. A federal government program used by all research agencies to develop new technologies.

SCS - Soil Conservation Service

SGI - Silicon Graphics Incorporated

Spatial data - data with implicit or explicit information about location such as latitude and longitude

TM - Thematic Mapper, a space-borne, multi-spectral imaging device.

USGS - United States Geological Survey. Part of the Dept. of Interior.

UTM - Universal Transverse Mercator. A map projection.

VAT - Visual Analysis Team. We exist but to serve.

Vector - Data are stored as explicit strings of x,y. From this simple definition there are many wildly different implementations. Several of the more common or significant examples are:

1. whole polygon structure - each polygon stored separately shared boundaries are stored more than once.

2. DIME - Dual Independent Map Encoding - developed by U.S. Bureau of Census and used as an archival and data exchange format. Each line segment stored with attributes. A major data source

3. Arc-Node - hierarchical - nodes are stored, then arcs form node to node and then polygons which are combinations of arcs. Data attributes stored with topographic data.

4. Relational structure - arc node structure with attributes stored separately in relational tables.

5. Digital Line Graph Structure (DLG) - USGS (U.S. Geological Survey) format. Data are subdivided into thematic layers. 1st layer - boundary info, 2nd - hydrographic features, 3rd - transportation network for area, 4th - public land survey systems. Features are broken down by codes. A major data source.

6. TIGER - Topologically Integrated Geographical Encoding and Reference system. Used for 1990 US census. A major data source.

APPENDIX C - Bibliography

Clark, D.M., D.A. Hastings and J.J. Kineman (1991), p 228, Global Databases and Their Implications for GIS.

Jeffrey Star and John Estes, 1990. Geographical Information Systems (An Introduction). New Jersey: Prentice-Hall, Inc.

Corporate Descriptions. Photogrammetric Engineering and Remote Sensing. Volume LVIII, No. 5, p. 489-760.

Maguire, David J., Michael F. Goodchild, and David W. Rhind (eds) 1991. Geographical Information Systems: Principle and Applications. Essex, England: Longman Scientific and Technical.

Marble, D.F.(1987). Geographical Information Systems: An Overview. In GIS for Resource Management: A Compendium, W.J. Ripple (ed.). Falls Church, VA: Amer. Soc. for Photogrammetry and Remote Sensing, pp. 2-8.

APPENDIX D - Vendor Data

The following is included to provide the reader with a baseline for costs. There is no implication that any one package is preferable. It is probable that several can largely satisfy most immediate requirements. Please note that the vendors have many versions and subsets of their products. We have attempted to show probable or at least reasonable selections. The guidelines we have used are:

hardware - What is the required hardware? If not SGI or PC, the vendor must specify, and the cost of same stated. If the software runs on either SGI or PC platforms, we assume that the government will spend approximately 2K to obtain larger hard disks and for the PCs approximately 5K for an Imagraph card and a 17-inch monitor for image display.

software - The software must be well designed to handle raster files with minor requirements for vector data. Specific minimal capabilities must include: tape handling, image display with vector overlays, geometric corrections, cell size resampling, macro language, classification, analysis between multiple layers of raster data, selection of raster areas based on vector boundaries, slope/ aspect and statistical analysis (frequency and distribution tools, correlation and covariance, and principle components at least, auto-correlation, factor analysis at the upper end). We must be able to add and/or modify the code.

There are numerous other things the software will have to do but this covers most of the rudimentary requirements which will apply to virtually all user's needs.

There are also multiple packages which are in the public domain. These include GRASS, ELAS, and LAS. Their procurement costs are nearly zero. However, they are not supported as are the commercial packages.

ARC/INFO Environmental Systems

Research Institute,Inc. (ESRI), 380 New York Street, Redlands, CA 92373

(714)793-2853 contact: Jorg Land (ext. 1118)

For the PC, there is a product called PC ARC/INFO for 4K per machine

For SGI:

basic package 7.10K * 5 = 35.5K
GRID (raster option) 1.25K * 5 = 6.25K
TIN (to meet min. req.) 1.25K * 5 = 6.25K support (1 lic.) 1.47K = 1.47K (lic. 2-5) 0.74K * 4 = 2.96K
ARCSDL (object code) 26.0K = 26.0K
Support ARCSDL 1.10K = 1.10K

Total 79.53K

Notes: This is the largest, in terms of installed customer base, of the large vector-based packages. It is included here for comparison. We do not believe that such a package is necessary for the Division, given current or near-term needs.

Delta Data Systems

contact Ren Clark 601-799-1813

Hardware: SGI or PC with Imagraph

Software: PC based - 5 copies * 6K = $30K. SGI based - 5 copies * 12.5K = 62.5K. These are 50% quantity discounted and each includes all software components of the complete AGIS package.

Notes: This is a third-generation software package. Unfortunately it does not use one of the standard GUIs for most of the product line. It also suffers from poor documentation. However, they have some of the most powerful and innovative code in the industry. A low-cost system is available which runs under MicroSoft Windows. AGIS is also one of the commercial spin-offs from ELAS.

Desktop Mapping System

contact Roy Welch 706-542-2359

Hardware: PC based using super-VGA

Software: $4950 1st copy 4*0.75*4950 for copies 2-5.

Total = $19.8K

Notes: This is a fairly simple system with good functionality. It does not have significant statistical or terrain analysis functions. It does not have a macro language. What it can do, it generally has only one way of doing; therefore, there is relatively little redundancy or flexibility.

ERDAS, Inc.

135 South Main Street, Box 31, Greenville, SC 29601

contact Paul Beaty (803)242-6109, (803)370-3908 fax

For the PC there is PC ERDAS which runs between 6K and 9K.

For the SGI (25% off for 2nd-5th, and 50% off for 6th and above)

Basic system - Imagine 8K + 24K = 32K C-programmers tool kit 4K + 12K = 16K

Total price for 5 copies = 48K

Options: Spatial modeler 4K + 12K = 16K software support 3K + 9K = 12K per year

Total for package = $76K

Notes: The raster GIS with probably the largest share of the middle price range market. This is another commercial package that is a spin-off from ELAS.

ER-Mapper

contact Eric Augustine 619-558-4709

Hardware: Currently SUN based, SGI version expected in June/July. Approximate cost for a suitable SUN (24-bit display with 1.2 GByte drive) is 8-9 K.

Software: Complete system is sold as a whole. There are no independent units. The government cost for 5 licenses would be 5*19.5*0.75*0.85= $62.2K. This is five copies at list price times quantity discount times government discount. Support and maintenance is an extra $5.8K

Notes: This is a third-generation software system produced by an Australian company. The software is native to X windows; it is not a port or a GUI on top of a command line interface. Source code is not available but users may add code.

Intergraph Corp.

Huntsville, AL 35894-0001

contact John Smith 205-430-5364 or (205)730-2000

PC software sold separately by Intergraph runs approximately 1-3K with specific hardware upgrades available.

The bulk of Intergraph software runs on their workstations which cost between 10K and 30K depending on options. The software pricing is variable and would depend on options selected.

Notes: A high end CAD/GIS running on sole source hardware, now migrating into compliance with industry standards.

MicroImages, Inc.

201 North 8th Street, Suite 15, Lincoln, NE 68508-1347

contact Lee D. Miller (402)477-9554, (402)477-9559 fax

The PC version costs between 4-6K depending on type of graphics resolution. The SGI version is 10K for single user and 30K for up to eight users accessing the software from one machine. For our assumed conditions our cost would be at least $20 - 30K for PC-based operations and $50K for SGI-based operations. There is a software development kit available for 3K.

A local contact (user) is Harold Pirtle (UAH professor) phone number: 776-2478

PCI

contact Mary-Viv Lawson
2221 Peach Tree Rd., NE, Suite D 216, Atlanta, GA 30309
(404)377-2002, fax (404)377-0906

Hardware: SGI or PC with Imagraph software: would be done as two licenses, which would run on six machines. To obtain source code with executables, double the cost of the individual components.

Software: Cost for five "seats"

basic package 12K
tape I/O 2.75K
multilayer analysis 2.75K
terrain analysis 2.75K
radar analysis 6K FFT 2.75K
atmospheric Correction 2.75K
programmers tool kit 2.75K
support 2.75K

Total Software Cost 37.25K

Notes: A full-featured system. A Canadian company.

Terra-Mar

1937 Landings Drive, Mountain View, CA 94043

contact David Butts (415) 964-6900

For the PC software a total package (including hardware display card update) and MicroImage software will run 11K.

For the SGI workstations, the license on a server costs 18K and each "seat" would cost 3.7K each. The total cost for five seats with the software tool kit (3.5K) would be 40K.

APPROVAL

A SURVEY OF GEOGRAPHICAL INFORMATION SYSTEMS APPLICATIONS FOR THE EARTH SCIENCE AND APPLICATIONS DIVISION, SPACE SCIENCES LABORATORY, MARSHALL SPACE FLIGHT CETNER

By

D. Rickman, K. A. Butler, and C. A. Laymon

This report has been reviewed for technical accuracy and contains no information concerning national security or nuclear energy activities or programs. The report, in its entirety, is unclassified.

Gregory S. Wilson
Director, Space Sciences Laboratory

☆U.S. GOVERNMENT PRINTING OFFICE 1994—533108/00040